

Genassemblage software, a tool for management of genetic diversity in human-dependent populations

Dariusz Kaczmarczyk

Received: 12 December 2013 / Accepted: 30 September 2014 / Published online: 10 October 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Genassemblage 1.0 is a user-friendly, Windows-based program that facilitates the assemblage of mating pairs to maintain genetic diversity. The software can be used with diploid, tetraploid and partially tetraploid organisms. To identify optimal mating pairs, the program estimates and compares the expected values of three genetic variation indicators for the progeny of individual spawning pairs (heterozygosity, percent of “weak heterozygotic” genotypes and allelic diversity). The genetic profiles of the potential parents include their tag numbers, sex and a list of alleles detected at investigated loci. The results are presented in tables, and the pairings that would have progeny with the best indicator values are highlighted; these values can be weighted to determine the best set of breeding pairs. Genassemblage files can be converted to .arp and .dat format for subsequent analysis with MSA and Arlequin 3.5.

Keywords Computer program · Genetic variation · Genetic profiles · Assemblage of mating pairs

Although it is important to maintain high levels of heterozygosity and allelic diversity in broodstock (Lande and Shannon 1996; Hansen et al. 2000), genetic variation within hatcheries-dependant populations may be decreasing due to progressive elimination of allelic diversity (Koljonen et al. 1999, 2002; Verspoor et al. 2005). Transfer of genetic variation to the next generation will be reduced if genetically similar individuals are paired because their

genetic profiles are unknown. To create breeding pairs who are as genetically different as possible, genetic profiles can be prepared using highly polymorphic fragments of DNA such as microsatellites. Unfortunately, the calculations involved are time-consuming, especially with animals with a tetrasomic genome, and there has been no software available for this task. To address this need, Genassemblage was constructed. It can be used for animals with diploid or tetraploid genomes, and animals with a diploid genome containing tetraploid fragments, e.g. cyprinids, salmonids and sturgeons.

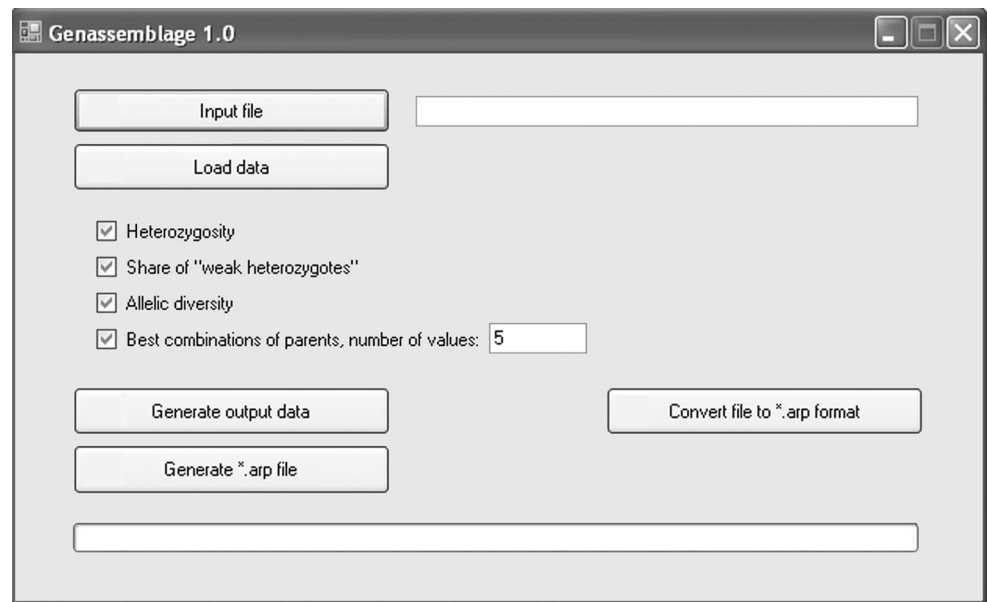
Genassemblage estimates the expected genetic variation of the offspring of different combinations of individuals. It calculates expected values for the heterozygosity of the potential offspring of each pairing, the number of different alleles the offspring would inherit, and, at tetrasomic loci, the percentage of “weak heterozygotes” [individuals with three identical alleles and a fourth allele that differs within a tetrasomic locus, e.g. AAAB (Kaczmarczyk and Fopp-Bayat 2012)]. To identify the best combinations, the values can be weighted based on their importance, then used to calculate a “ v index” indicating the best combinations.

Optimal breeding pairs are selected based on the following criteria and assumptions:

- A. The sex of the individuals is known; each individual is fertile and ready for reproduction.
- B. All individuals are marked and identifiable.
- C. The probability of two identical genotypes in unrelated individuals is very small.
- D. Individuals with a higher frequency of heterozygous genotypes at the analysed genetic markers are more likely to have heterozygous genotypes in loci that are important for viability and adaptation (Shikano and Taniguchi 2002; Olech 2003).

D. Kaczmarczyk (✉)
Department of Environmental Biotechnology, University of
Warmia and Mazury in Olsztyn, Słoneczna 45G,
10-718 Olsztyn, Poland
e-mail: d.kaczmarczyk@uwm.edu.pl

Fig. 1 The Genassemblage user-interface



- E. If individuals have different alleles within the analysed loci, they are more likely to differ in loci that determine viability and adaptation than individuals with identical alleles.

Genassemblage 1.0 is a Windows-based program; it can be downloaded with a detailed user manual from http://pracownicy.uwm.edu.pl/d.kaczmarczyk/main_page.htm. The program can convert *.xls or *.dat files to *.arp. files, so data can be transferred from Genassemblage to Arlequin 3.5 (Excoffier and Lischer 2010), and to MSA (Dieringer and Schlötterer 2003).

Figure 1 shows the user interface. It allows the user to load data, choose the values that Genassemblage will calculate, and convert files to .arp format. The input data are the genetic profiles of the investigated individuals. Each profile includes the individual's population group and population name, its tag number and sex, and a list of the alleles of investigated markers detected in its genome. These markers should have the following characteristics:

- A. a high level of polymorphism,
- B. inheritance on autosomal chromosomes following Mendelian Laws,
- C. neutrality for evolution (not subject to natural selection).

The expected heterozygosity of the offspring is calculated from disomic and tetrasomic markers for which there are complete genotyping results for both potential parents. A locus for which there is no data from one or both parental individuals is excluded from the calculations for this breeding pair. The following equation is used

$$H = 1 - \frac{(ph_1 + ph_2 + ph_3 + \dots ph_n)}{nl} \quad (\text{Algorithm1})$$

Algorithm 1. Calculation of the expected heterozygosity of offspring in which ph_n is the expected share of homozygous genotypes in the offspring of a specific breeding couple within the n th loci, and nl is the total number of analysed loci.

To calculate the number of different alleles that potential offspring would inherit, all the alleles found in the markers (na_n) are summed (Algorithm 2). The calculations include all the alleles found within the di- and tetrasomic markers in a specific breeding couple.

$$ar = \sum (na_n) \quad (\text{Algorithm2})$$

Algorithm 2. Calculation of potential allelic diversity in the offspring of a specific couple.

The expected proportion of weak heterozygote loci in offspring is calculated only for tetrasomic markers using Algorithm 3, where wh is the probability that the locus n is a "weak heterozygote. The values of probability (pwh) are calculated assuming that alleles of the tetrasomic fragment are located on four independently inherited homologous chromosomes, gametes conjugate randomly, and the frequency of individual genotypes in the offspring is not changed by natural selection.

$$wh = \frac{(pwh_1 + pwh_2 + pwh_3 + \dots pwh_n)}{nl} \quad (\text{Algorithm3})$$

Algorithm 3. Calculation of the expected proportion of "weak heterozygotes".

Figure 2 shows how the results of the above calculations are displayed. For estimated heterozygosity and allelic

Fig. 2 The expected heterozygosity, allelic diversity and percentage of “weak heterozygotes” in progeny of potential spawning pairs. *Yellow* (here, *light grey*), indicates incomplete data for one parent. *Orange* (here, *dark grey*) indicates incomplete data for both parents

ESTIMATED HETEROZYGOSITY IN PROGENY OF DIFFERENT COMBINATIONS OF SPAWNERS								
		FEMALE						
		A06	A07	A14	B19	B34	C38	C40
MALE	A01	0.4931	0.6667	0.6042	0.9236	0.6667	0.6042	0.8611
	B05	0.8750	0.8681	0.6875	0.7708	0.8125	0.8750	0.9167
	B20	0.6389	0.6181	0.7083	0.9028	0.6458	0.6806	0.9630
	C41	0.6667	0.8056	0.7500	0.9167	0.7500	0.8056	0.9583
	C42	0.7500	0.8542	0.7500	0.9167	0.8125	0.8542	0.9722
The share of “weak heterozygotes” in progeny of different combinations of spawners (all = 1)								
		FEMALE						
		A06	A07	A14	B19	B34	C38	C40
MALE	A01	0.4097	0.3403	0.4583	0.4722	0.5000	0.2778	0.4861
	B05	0.5000	0.4028	0.4792	0.5000	0.3750	0.0417	0.2917
	B20	0.4931	0.1944	0.4167	0.4722	0.5000	0.1667	0.3750
	C41	0.5000	0.4722	0.5833	0.5833	0.5833	0.5139	0.5139
	C42	0.3750	0.3611	0.5417	0.5139	0.4167	0.2986	0.3819
The number of alleles inherited by progeny of different combinations of spawners								
		FEMALE						
		A06	A07	A14	B19	B34	C38	C40
MALE	A01	8	9	9	11	10	10	9
	B05	12	12	10	10	10	14	11
	B20	9	9	11	9	8	11	10
	C41	8	9	9	10	9	8	8
	C42	12	13	12	13	12	12	11

diversity; higher values are better; to reduce the number of homozygotic individuals in the following generations, a lower number of “weak heterozygotes” is better. These values can be weighted to calculate a “v index” (see the Genassemblage manual) that indicates the best parental combinations.

Genassemblage can also be used for other tasks related to managing genetic variation in breeding stocks or human dependant populations.

Acknowledgments I would like thank Dr Pawel Brzuzan and Dr. Mirosław Luczynski for their help in beta testing this software, and for all the helpful suggestions which were incorporated in the preparation of this software and its manual. The creation of Genassemblage software has been supported by the Polish Ministry of Science (Project Number N N304 400338).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Dieringer D, Schlötterer C (2003) Microsatellite analyzer (MSA): a platform-independent analysis tool for large microsatellite data sets. *Ecol Notes* 2:1–3. doi:[10.1046/j.1471-8286.2003.00351.x](https://doi.org/10.1046/j.1471-8286.2003.00351.x)
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows”. *Mol Ecol Resour* 10:564–567. doi:[10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x)
- Hansen MM, Nielsen EE, Ruzzante DE, Bouza C, Mensberg K (2000) Genetic monitoring of supportive breeding in brown trout (*Salmo trutta* L.), using microsatellite DNA markers. *Can J Fish Aquat Sci* 57:2130–2139. doi:[10.1139/cjfas-57-10-2130](https://doi.org/10.1139/cjfas-57-10-2130)
- Kaczmarczyk D, Fopp-Bayat D (2012) Assemblage of spawning pairs based on their individual genetic profiles—as a tool for maintaining genetic variation within sturgeon populations. *Aquac Res* 44:77–682. doi:[10.1111/j.1365-2109.2011.03064.x](https://doi.org/10.1111/j.1365-2109.2011.03064.x)
- Koljonen M-L, Jansson H, Paaver T, Vasin O, Koskiniemi J (1999) Phylogeographic lineages and differentiation pattern of Atlantic salmon in the Baltic Sea with management implications. *Can J Fish Aquat Sci* 56:1766–1780. doi:[10.1139/f99-104](https://doi.org/10.1139/f99-104)
- Koljonen M-L, Tähtinen J, Säisä M, Koskiniemi J (2002) Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programs and the geographic distribution of microsatellite variation. *Aquaculture* 212:69–92. doi:[10.1016/S0044-8486\(01\)00808-0](https://doi.org/10.1016/S0044-8486(01)00808-0)
- Lande R, Shannon S (1996) The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50:434–437. doi:[10.2307/2410812](https://doi.org/10.2307/2410812)
- Olech W (2003) The influence of individual and maternal inbreeding on European bison (*Bison bonasus*) calves survival. Treatises and monographs. Wydawnictwo SGGW, Warszawa, Poland (in Polish with English summary)
- Shikano T, Taniguchi N (2002) Relationships between genetic variation measured by microsatellite DNA markers and a fitness-related trait in the guppy (*Poecilia reticulata*). *Aquaculture* 209:77–90. doi:[10.1016/S0044-8486\(01\)00812-2](https://doi.org/10.1016/S0044-8486(01)00812-2)
- Verspoor E, Beardmore JA, Consuegra S, García de Léniz C, Hindar K, Jordan WC, Koljonen M-L, Mahkrov AA, Paaver T, Sánchez JA, Skaala Ø, Titov S, Cross TF (2005) Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *J Fish Biol* 67:3–55. doi:[10.1111/j.0022-1112.2005.00838.x](https://doi.org/10.1111/j.0022-1112.2005.00838.x)